ECOGRAPHY

Research

Evolutionary patterns in the geographic range size of Atlantic Forest plants

Tarciso C. C. Leão, Eimear Nic Lughadha and Peter B. Reich

T. C. C. Leão (https://orcid.org/0000-0002-7029-1582) ✉ (t.leao@kew.org) and P. B. Reich, Dept of Forest Resources, Univ. of Minnesota, St Paul, MN, USA. PBR also at: Hawkesbury Inst. for the Environment, Western Sydney Univ., Penrith, NSW, Australia. – E. Nic Lughadha and TCCL, Royal Botanic Gardens, Kew, Richmond, UK. TCCL also at: CAPES Foundation, Ministry of Education, Brasilia, Brazil.

Ecography
43: 1510–1520, 2020
doi: 10.1111/ecog.05160

Subject Editor: Wolf Eiserhardt
Editor-in-Chief: Miguel Araújo
Accepted 24 June 2020

Species' geographic range size is arguably the single most important predictor of vulnerability to extinction and a key metric in ecology. Despite this, patterns of specific variation in range size and their underlying reasons are still poorly understood. For example, hypotheses on how evolutionary history affects range size have scarcely been tested. To address these questions, we focused on Brazil's Atlantic Forest flora, one of the most species-rich in the world, relatively well-known and highly threatened. We investigated whether and how lineages' diversification rate, number of species and age are associated with species' geographic range size. We estimated the extent of occurrence and area of occupancy of each of 13 283 plant species native to the Atlantic Forest region based on over 500 000 unique records. We used phylogenetic least squares and logistic regressions to analyze how the predictors affect the geographic range size. On average, the higher the diversification rate and number of species in the lineage, the smaller the species range size and the higher the proportion of species with vulnerably small range size. Lineage age showed no clear effect on average range size. The results support our expectations that dynamics of diversification and taxonomic richness considerably affect the species range size. Finally, this work reveals poorly known patterns of range size variation and some of the mechanisms driving variation in range size and vulnerability to extinction.

Keywords: area of occupancy, diversification rate, evolutionary age, extent of occurrence, species richness, tropical forest, vulnerability to extinction

Introduction

The geographic range size is a powerful indicator of the area over which a species is distributed and the species' vulnerability to extinction (Gaston 2003, Leão et al. 2014). Because of this, it is a key metric in the fields of ecology and conservation biology. The distribution of geographic range sizes varies by many orders of magnitude across regions of the world, taxonomic groups and life forms (Brown et al. 1996, Gaston 2003). Although there is consistent evidence of range size variation being associated with species-specific traits among groups of animals (e.g. body size patterns), patterns of range size variation among plants are much less known and consistent (Murray et al.

2002), especially among tropical plants. Perhaps the best known pattern in plant range size is associated with the available area of habitat; the larger the available area of habitat the larger the range size (Morueta-Holme et al. 2013, Gallagher 2016). Other emerging, but generally poorly tested, patterns of range size are associated with the species evolutionary history.

The evolutionary history of plant lineages seems to explain a considerable amount of variation in species' geographic range size (Schwartz and Simberloff 2001, Lozano and Schwartz 2005, Paul et al. 2009) and extinction risk (Davies et al. 2011). Plant taxa with more species have higher proportions of rare species, which suggests that rates of net diversification and/or patterns of taxonomic delimitation affect species' range sizes (Schwartz and Simberloff 2001, Lozano and Schwartz 2005). Old species of neotropical *Psychotria* tend to occupy a larger proportion of their extent of occurrence than young species (Paul et al. 2009), consistent with an old hypothesis of a positive relationship between age and area (Willis 1922). Young and fast-evolving plant lineages have higher extinction risk in the Cape region of South Africa, probably due to naturally small species' geographic range sizes (Davies et al. 2011), though this is not always the case (Tanentzap et al. 2020). These findings support the existence of a biologically relevant effect of evolutionary history on geographic range sizes of extant plant species.

A key to understanding how evolutionary history affects geographic range size likely rests on the dynamics of diversification. A speciation event creates descendant species that by definition initially have smaller average range size than the ancestral species. Even if a descendant species has a range size that is as large as that of its ancestor, the mean range size of all descendant species must have been, at the start, smaller than the range size of the ancestor. This fact suggests that, all else being equal, clades with higher speciation rate should have species with smaller average range sizes. Given that species with smaller range sizes are more vulnerable to extinction (Rosenzweig 2001, Gaston 2003, Gaston and Fuller 2009), as the proportion of small-ranged species in a lineage increases, the extinction rate in that lineage is likely to increase as well (Rosenzweig 2001, Castiglione et al. 2017). If the speciation rate largely exceeds the extinction rate, then there should be an accumulation of small-ranged species in the lineage. Thus, a high ratio between speciation and extinction is expected to lead both to an elevated number of species and an increased proportion of small-ranged species (Schwartz and Simberloff 2001, Lozano and Schwartz 2005). This rationale sets our expectations for how diversification rate should affect current species range size and vulnerability to extinction. Alternatively, if relatively young, recently diverged species with small range sizes are expanding their range, they may not be at similar risk of extinction as older species with similarly small ranges that may have shrunk over time.

Geographic range size is probably the single most relevant factor affecting species' extinction risk (Leão et al. 2014, Nic Lughadha et al. 2019). A species that occurs in a small area is vulnerable to disappearing completely due to habitat loss and environmental change (Gaston 2003, Gaston and Fuller 2009). In fact, critically small range size as defined under criterion B1 (extent of occurrence, hereafter EOO) or criterion B2 (area of occupancy, hereafter AOO) is the main reason cited for categorization of species as threatened in the IUCN Red List of Threatened Species (hereafter Red List, Gaston and Fuller 2009, Brummitt et al. 2015). A full account of extinction risk also depends on factors such as population size and trend, and on disturbances causing declines in population size or range size (IUCN 2012), but most, if not all, species with declining populations will ultimately show small range size before extinction, highlighting the importance of range size as a predictor of extinction risk. Thus, understanding the patterns and mechanisms of range size variation is key to understand species' vulnerability to extinction.

With more than 33 000 known angiosperm species, Brazil is the country with the largest documented plant diversity in the world, and the Atlantic Forest has the largest known plant diversity in Brazil, making it one of the richest known floras in the world (Zappi et al. 2015, The Brazil Flora Group et al. 2018). The largely fragmented and disturbed Atlantic Forest (Ribeiro et al. 2009) has 15 000 known flowering plant species (The Brazil Flora Group et al. 2018), which represents almost half of Brazil's known angiosperm flora and 60% of Brazil's known threatened plant species (Martinelli and Moraes 2013, Martins et al. 2018). The Atlantic Forest region is also home to most of Brazil's human population and economic activity (Scarano and Ceotto 2015). The combination of extraordinarily high plant diversity with high human pressure makes the Atlantic Forest a region where considerable numbers of plant species are likely to go extinct in this century. Beyond its high conservation concern, the Atlantic Forest currently provides a unique opportunity to investigate the existence of patterns in the range size of tropical plants, as it is likely among the best collected tropical floras on a large geographical scale, with millions of herbarium records digitally available (Species Link Network, inct.splink.org.br).

Previous research investigating predictors of extinction risk in the Atlantic Forest found that closely related species tend to have similar geographic range size, although the causal factors remain unclear (Leão et al. 2014). In this study, we build a new dataset including evolutionary attributes, more species and species occurrences, and an improved phylogeny to investigate evolutionary patterns in the variation of species' geographic range size. Focusing on the Atlantic Forest's flowering plants, we investigated the effects of lineage diversification rate, species richness and evolutionary age on the species' geographic range size and vulnerability to extinction. Specifically, we tested whether species of younger clades, and of clades with rapid diversification and/or higher species richness have on average smaller geographic range size and higher vulnerability to extinction (Schwartz and Simberloff 2001, Lozano and Schwartz 2005, Paul et al. 2009, Davies et al. 2011).

Methods

Species occurrences

We downloaded 3 785 741 raw occurrences of plants from more than 100 herbaria and digital collections through the Species Link Network (<www.splink.org.br>, see section '1. List of Herbaria providing data' in the Supplementary material Appendix 1). After cleaning the data (see Cleaning steps below), filtering only native species occurring in the Atlantic Forest (according to Brazilian Flora 2020, reflora.jbrj.gov.br) and keeping only one species occurrence per administrative unit regardless of the collection date, 502 000 occurrences representing 13 283 species remained and were analyzed in this study (i.e. ~90% of all native angiosperms documented for the Atlantic Forest). A single occurrence may encompass multiple herbarium records collected in the same administrative unit on different dates or different locations within the administrative unit. The range in collection dates per species had mean of 22 yr and maximum of 379 yr. We used the centroid of the most local administrative unit (i.e. district or municipality) as the reference coordinate for each occurrence. Using the administrative unit instead of specific coordinates allowed us to include many more occurrences and to avoid errors, because geographic coordinates were often not available or not reliable whereas administrative units were usually clearly stated. The administrative units are relatively small across the Atlantic Forest region, where most of the occurrences are concentrated, with a median distance between closest administrative units of 10 km (see details in section 2 of the Supplementary material Appendix 1). The limits of municipalities and districts were obtained from the database of Global Administrative Areas ver. 2.5 (<www.gadm.org>).

Cleaning steps

We conducted several cleaning and standardization procedures on the occurrence data before using it for analysis. All unusual characters in the species binomials were removed (e.g. '?', '_', '.', '#', '*', etc.). All name strings reflecting uncertain identification were removed (i.e. those including 'cf.', 'aff.' or at genus level 'sp.', 'spp.'). Infraspecific taxa (e.g. subspecies and varieties) were grouped at the species level. Synonyms were checked and accepted names were adopted according to Brazilian Flora 2020 (reflora.jbrj.gov.br, Brazil Flora Group 2019). Algorithms for matching synonyms and adjusting minor spelling errors in the raw binomials were applied during the synonym checking process through the function 'get.taxa' from the package 'flora' (Carvalho 2016), which match names to the Brazilian Flora 2020 taxonomy (Brazil Flora Group 2019). No spelling corrections were allowed on genus name, and only minor corrections were allowed on specific epithet to ensure correct reconciliation of misspelled binomials. All records that could not be matched with an accepted name at Brazilian Flora 2020 (Brazil Flora Group 2019) were discarded. Discarded names were mainly illegitimate, invalid or rejected names.

Records with no administrative unit and with coordinates that were likely wrong or lacking coordinates (e.g. positioned in the ocean, coordinates not matching reported administrative unit, and coordinates equal zero) were removed. We kept only occurrences within South American boundaries, which helped to avoid the inclusion of potentially introduced occurrences and had only minor impact, understating the range of few widespread species (2% of the species had occurrences outside South America).

Estimation of geographic range size

The species area of occupancy (AOO) was the sum of the area of grid cells occupied by the species. Grid cell size varied per species according to our adaptation of the 'sliding scale' method proposed by Willis et al. (2003). This method allows assignment of a cell size appropriate to the distribution extent of each species, avoiding misrepresentations of their AOO. The width of the grid cell for a species was defined as one tenth of the maximum distance between occurrences. A minimum (5 × 5 km) and maximum (50 × 50 km) cell size was set to avoid undesired distortions, such as the AOO being too small when there are too few occurrences available, or the area being too large when there are few occurrences with very sparse distribution. The AOO of each species was calculated in R with the support of the packages 'raster' (Hijmans 2016), 'sp' (Bivand et al. 2013) and 'rgdal' (Bivand et al. 2016).

The extent of occurrence (EOO) was defined as the area in square kilometers of the smallest convex hull that enclosed all occurrences of the species, as computed by the quickhull algorithm (Barber et al. 1996) through function 'convhulln()' of the package 'geometry' (Habel et al. 2015). The EOO was calculated only for species occurring in three or more administrative units.

A species AOO smaller than 2000 km^2 and EOO smaller than 20 000 km^2 was considered 'vulnerably small' in accordance with the criteria B1 and B2 of the IUCN Red List (IUCN 2012). The classification based on Red List thresholds allows the investigation of patterns in species extinction risk. Although a full account of the vulnerability to extinction should consider other factors beyond range size, very small range size is a good indicator of vulnerability to extinction (Gaston 2003, Gaston and Fuller 2009).

Predictor variables

We defined clade as the smallest phylogenetic group that includes all Atlantic Forest species of a genus according to our reference phylogeny 'ALLMB' (Smith and Brown 2018). The 'ALLMB' phylogeny includes 356 305 terminal taxa from GenBank and Open Tree of Life with the backbone from Magallón et al. (2015). We estimated the net diversification rate of each clade's crown-group based on the birth–death model of diversification when extinction is either negligible ('pure-birth,' $\epsilon = 0$) or high ($\epsilon = 0.9$) (Magallon and Sanderson 2001, Harmon 2019). These are simplistic models that enable estimation of net diversification rates for clades

with poorly defined internal nodes by using the number of extant species and age of the most recent common ancestor (MRCA). Net diversification rates were estimated for each clade's crown-group according to the equation (Magallon and Sanderson 2001, Harmon 2019):

$$\hat{r} = \frac{\ln\left[\frac{n\left(1-\epsilon^2\right)}{2} + 2\epsilon + \frac{(1-\epsilon)\sqrt{n\left(n\epsilon^2 - 8\epsilon + 2n\epsilon + n\right)}}{2}\right] - \ln(2)}{t}$$

where $\hat{r}$ = predicted net diversification rate, n = number of extant species in the clade, t = age of clade's crown-group MRCA in million years, ϵ = relative extinction rate. Monotypic genera were forced to have zero net diversification rate to avoid negative values in the log-transformation used in the regression models. Age of the MRCA is the estimated time (in millions of years) from the present to the node of the most recent common ancestor that includes all extant Atlantic Forest species of the genus. Age estimates and number of extant species in the clade were obtained from the full dated phylogeny 'ALLMB' (Smith and Brown 2018).

We also estimated a species-level age, as the length of the tip (or divergence time) in millions of years (My) using the full 'ALLMB' phylogeny, but analyzed species age only for those species with genetic data available (i.e. 3676 species and 1366 genera). It is worth noting that the species with genetic data available have on average larger range size than the full set of Atlantic Forest species (50% larger mean EOO and 70% larger mean AOO), probably because widespread species were more likely to be sampled in genetic studies. We processed and edited phylogenetic trees with support of the R packages 'ape' (Paradis et al. 2004) and 'phytools' (Revell 2012), and used code from Title and Rabosky (2019) to estimate diversification rate.

We measured species richness as the number of accepted species of the genus that occur in the Atlantic Forest according to Brazilian Flora 2020 (reflora.jbrj.gov.br, Brazil Flora Group 2019). Species richness in the Atlantic Forest is more appropriate than global species richness per genus to investigate the effect of taxonomic richness on range size because it matches the sample of species used to measure the response variable (EOO and AOO of Atlantic Forest species). A mismatch between the samples used to measure predictor and response variables would make it hard to verify the effects of predictors on response variables (see details in section 4 of the Supplementary material Appendix 1).

Life forms, habitats and vegetation type were classified according to Brazilian Flora 2020 (reflora.jbrj.gov.br, Brazil Flora Group 2019). These are important species-level ecological variables that correlate with range size (Leão et al. 2014), thus they were included as covariates in the models to provide a more robust test of the effects of our focal predictors. Inferences regarding the effects of these ecological covariates are beyond the scope of this study. Details regarding these covariates are in section 3 of the Supplementary material Appendix 1.

Statistical analyses

We estimated and reported the effects of our predictor variables (i.e. net diversification rate, species richness and evolutionary age) on geographic range size while controlling for phylogenetic correlation and important covariates, namely life form, habitat and vegetation type. This approach provides a robust estimation of the independent effects of the predictors. We used phylogenetic generalized least squares regression (PGLS) to explore the effects of our predictors on continuous EOO and AOO, using Pagel's lambda (λ) as covariance parameter (Pagel 1999, Tung Ho and Ané 2014). $\lambda = 0$ suggests no phylogenetic signal, while $\lambda = 1$ suggests high phylogenetic signal equal to a Brownian motion model of evolution. We estimated the explained variance of models according to the R^2_{pred} method for phylogenetic models described by Ives (2018). To explore the effects on species extinction risk we used phylogenetic logistic regressions as described in Ives and Garland (2010) and implemented by Tung Ho and Ané (2014), using the method that maximizes the penalized likelihood of the logistic regression and alpha to estimate the phylogenetic correlation. The logistic regressions allow estimation of how the predictors affect the probability of a species having vulnerably small range size, which provides a more valuable indication of the species vulnerability to extinction than the regressions on continuous range size.

Clades where most species do not occur in the Atlantic Forest are expected to show weak relationships between diversification rate or species richness and species range size because diversification rate and species richness are properties of the entire clade or genus (in the Atlantic Forest), while the species range size is measured only for sampled species occurring in the Atlantic Forest (see detail in section 4 of the Supplementary material Appendix 1). A way to avoid this mismatch is to analyze only clades where all species occur in the Atlantic Forest, but this severely reduces sample size and representativeness of the Atlantic Forest. Because of this trade-off between reducing sample mismatch and increasing sample size, we performed analyses on ten subsets of clades with varying proportions of species occurring in the Atlantic Forest (from 0 to 100%) to allow visualisation of the consistency of results. We reported detailed results for three of those subsets of clades that are representative of the entire spectrum: 1) clades that have all of their species occurring in the Atlantic Forest (850 species, 6% of Atlantic Forest species); 2) clades that have at least half of their species occurring in the Atlantic Forest (1864 species, 12% of Atlantic Forest species); and 3) all clades with species occurring in the Atlantic Forest (9993 species, 67% of Atlantic Forest species). The number of species refers to those with complete data for analysis. Tables and figures are based on the subset of clades with at least half of the species occurring in the Atlantic Forest, which is a reasonable compromise between our aims to reduce sample mismatch and increase sample size.

We applied a systematic approach to fitting appropriate models, checking for non-constant variance and deviations from normality in the model residuals. We applied transformations that helped spread the data more symmetrically

and make it more appropriate to use in the regression modeling. We found the most appropriate transformations with the assistance of the 'symbox' function in the R package 'car' (Fox and Weisberg 2011). We transformed EOO, AOO, net diversification rate, species richness, MRCA age and species age to the logarithmic scale (base 2). Variable transformations helped to achieve linearity and better residual properties. Continuous variables were scaled and centred to allow for direct comparison of effect sizes.

Results

Species' AOO ranged from 25 km^2 to 1 500 000 km^2, with median of 27 000 km^2 and mean of 75 000 km^2. Species' EOO ranged from 8 km^2 to 11 800 000 km^2, with median of 576 000 km^2 and mean of 1 777 000 km^2 (Fig. 1a, b). To visualize the spatial variation in range size, we mapped the mean AOO and the standard deviation on 0.5° × 0.5° grid cells across the Atlantic Forest (Fig. 1c, d).

Evolutionary history and species' geographic range size

On average, the higher the lineage diversification rate and species richness, the smaller the species' range size (Table 1, Fig. 2a, b). A two-fold increase in the net diversification rate was associated on average with 16–17% (all clades: 8–9%; Atlantic Forest only: 16–17%) reduction in species EOO and 12–13% (all clades: 6–7%; Atlantic Forest only: 15–16%)

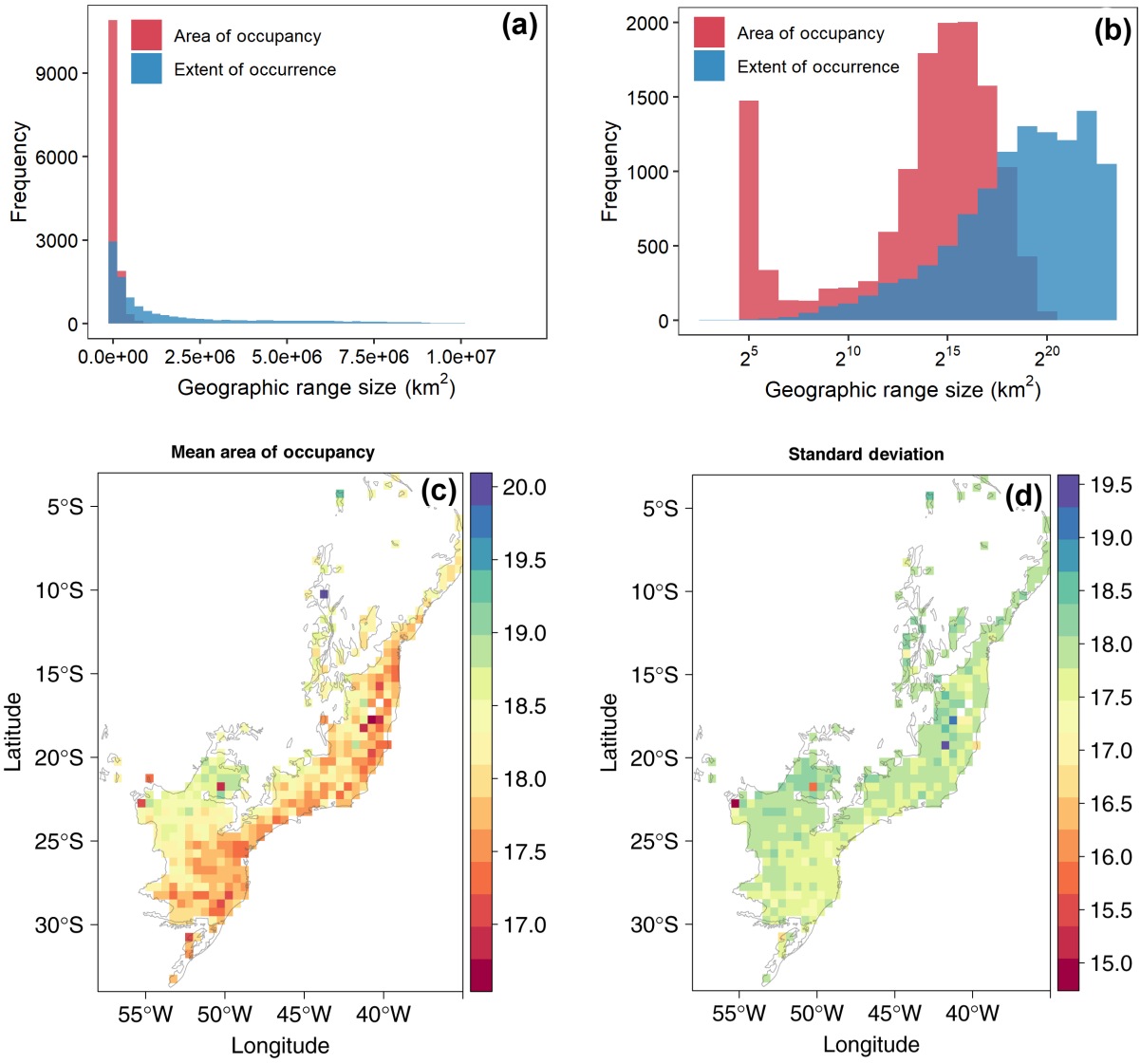

Figure 1. Frequency distribution of the (a) untransformed and (b) $\log_2$ transformed species geographic range size (area of occupancy and extent of occurrence) of Atlantic Forest angiosperms. Darker blue represents the overlap between the two distributions. Spatial distribution of $\log_2$ transformed species (c) mean area of occupancy and (d) standard deviation across the Atlantic Forest region summarized for 2852 grid cells with 0.5° longitude × 0.5° latitude resolution. Estimates of area of occupancy included 13 283 species, while estimates of extent of occurrence included 10 830 species (i.e. species with records in three or more administrative units).

reduction in species AOO. Variation is due to assumption of no extinction, $\epsilon = 0$, or high relative extinction rate, $\epsilon = 0.9$ (full spectrum of variation in Fig. 3). A two-fold increase in species richness was associated on average with 24% (all clades: 19%; Atlantic Forest only: 26%) reduction in species EOO and 19% (all clades: 15%; Atlantic Forest only: 25%) reduction in species AOO. These findings corroborate our expectations that species from lineages with higher diversification rates and larger number of species tend to have species with smaller geographic range size.

Age of the clade's most recent common ancestor (MRCA) showed weak negative average effects on geographic range size (Table 1), meaning that older clades were associated with slightly smaller average species range sizes. We found no significant effects of age when measuring at species level ('species age', i.e. length of terminal edge for species with genetic data; Supplementary material Appendix 1 Table A1).

We found considerable phylogenetic signal in EOO ($\lambda_{50\%} = 0.50, \lambda_{all} = 0.59, \lambda_{AF\text{-}only} = 0.59$) and AOO ($\lambda_{50\%} = 0.37, \lambda_{all} = 0.48, \lambda_{AF\text{-}only} = 0.34$), which was only partially correlated with the predictors. About 12 and 11% of the variance in EOO and AOO respectively was independently explained by the phylogeny after controlling for all predictors, while the full models explained 31 and 25% of the variance respectively (Table 1). The model residuals still showed significant phylogenetic signal for EOO ($\lambda_{50\%} = 0.35$–0.37, $\lambda_{all} = 0.54$–0.55, $\lambda_{AF\text{-}only} = 0.46$–$0.50$) and AOO ($\lambda_{50\%} = 0.24$–$0.26$, $\lambda_{all} = 0.43$, $\lambda_{AF\text{-}only} = 0.21$–$0.22$), meaning that inheritance from a common ancestor and/or other phylogeny-related variables may explain substantial variation in range size.

Evolutionary history and species' vulnerability to extinction

Twelve percent of the species had EOO smaller than the vulnerability threshold (i.e. < 20 000 km^2) and 20% had AOO smaller than the vulnerability threshold (i.e. < 2000 km^2). Five percent had both. A two-fold increase in the diversification rate was associated with 20–21% (all clades: 16%; Atlantic Forest only: 11–13%) increase and 19–21% (all clades: 11–16%; Atlantic Forest only: 15–20%) increase in the odds that a species has vulnerably small EOO or AOO, respectively (Table 2). Similarly, a two-fold increase in the species richness was associated with 23% (all clades: 20%; Atlantic Forest only: 22%) increase and 24% (all clades: 19%; Atlantic Forest only: 24%) increase in the odds that a species has vulnerably small EOO or AOO, respectively.

Table 1. Coefficients and statistical significance of the effects of clade net diversification rate (under assumptions of negligible extinction $\epsilon = 0$ or high extinction $\epsilon = 0.9$), species richness and MRCA age on species' extent of occurrence (n = 1490 species) and area of occupancy (n = 1864 species) according to phylogenetic generalized least squares regression. Data include all clades with at least 50% of their species occurring in the Atlantic Forest. Models included life form, habitat and vegetation type as covariates to enable for robust evaluation of the independent effects of variables of interest. All quantitative variables were log transformed, scaled and centred, allowing for direct comparison of effect sizes.

	Extent of occurrence			Area of occupancy		
Diversification rate $\epsilon = 0$	−0.15***			−0.09***		
Diversification rate $\epsilon = 0.9$		−0.14***			−0.09***	
Species richness			−0.23***			−0.14***
MRCA age			−0.06*			−0.05*
Life form						
Climber	0.15	0.11	0.19†	−0.01	−0.03	0.02
Herb	−0.15	−0.16	−0.12	−0.20*	−0.21*	−0.18†
Shrub	−0.40***	−0.41***	−0.35***	−0.35***	−0.35***	−0.31***
Tree						
Unknown/others	−0.37**	−0.40**	−0.35**	−0.31**	−0.32**	−0.28*
Habitat						
Aquatic	0.44*	0.47**	0.48**	0.37*	0.39*	0.39*
Epiphytic	−0.17	−0.18	−0.23†	0.00	−0.01	−0.04
Hemiepiphyte	0.28	0.32	0.27	0.36	0.38	0.34
Parasite	−0.28	−0.28	−0.20	−0.27	−0.27	−0.22
Rupicolous	−0.10	−0.09	−0.10	0.00	0.00	0.00
Saprophyte	0.29	0.28	0.29	0.25	0.24	0.23
Terricolous						
Unknown	0.46†	0.49†	0.49†	0.08	0.10	0.12
Vegetation type						
Open vegetation	−0.50***	−0.51***	−0.55***	−0.44***	−0.45***	−0.48***
Forest	−0.32***	−0.31***	−0.32***	−0.53***	−0.52***	−0.53***
Forest and open veg.						
Intercept	0.23	0.25	0.18	0.47**	0.48**	0.42**

†p < 0.1.
*p < 0.05.
**p < 0.01.
***p < 0.001.

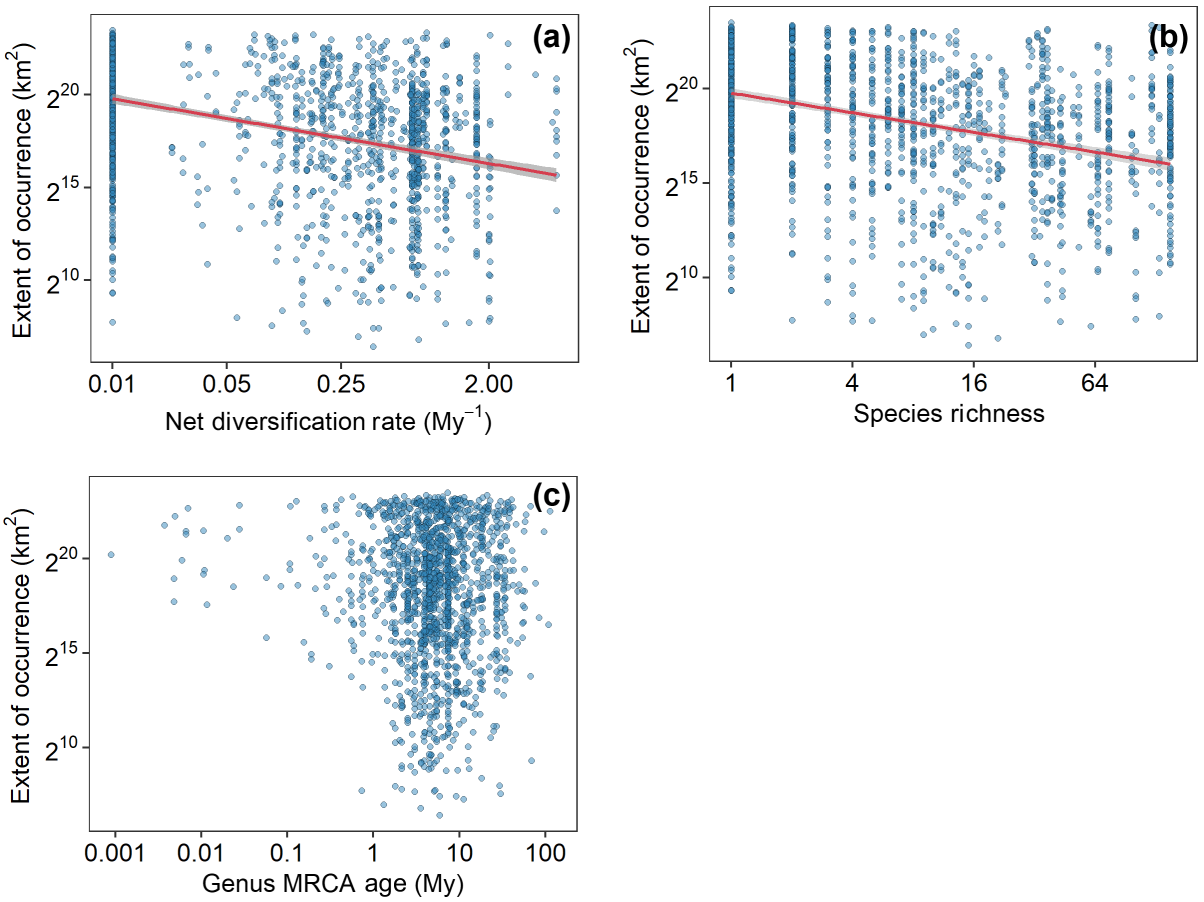

Figure 2. Bivariate relationships between species' extent of occurrence and (a) clade net diversification rate under assumption of no extinction ($\epsilon = 0$), (b) species richness, and (c) clade MRCA age, based on the subset of clades with at least half of their species occurring in the Atlantic Forest. Least squares regression line is shown for statistically significant relationships. Monotypic genera were forced to have net diversification rate of 0.01 to allow for log transformation. Shaded area shows the 95% confidence interval.

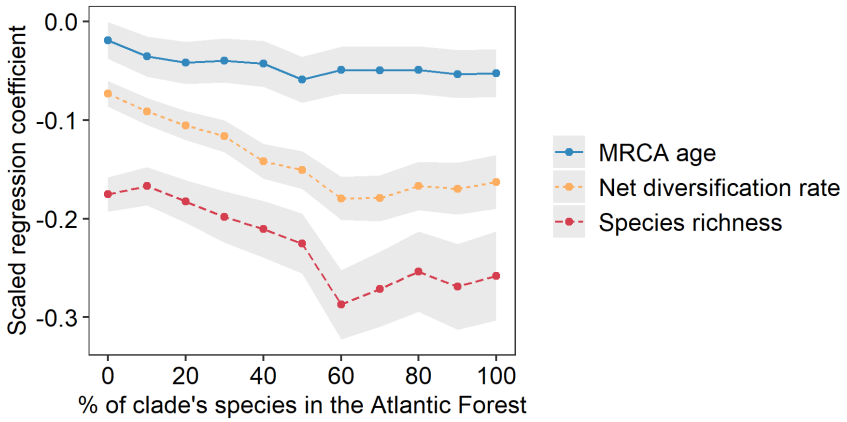

Figure 3. Coefficients of net diversification rate ($\epsilon = 0$), species richness and clade's MRCA as a function of percentage of species in the clade that occurs in the Atlantic Forest. Coefficients were estimated using phylogenetic generalized least squares models that include the same covariates as described in Table 1. Coefficients in the far left of the plot (i.e. ~0%) were estimated based on data from all clades (9993 species), and coefficients in the far right (i.e. 100%) were estimated based only on clades that have all of their species occurring in the Atlantic Forest (850 species). Variables were log transformed, scaled and centred, allowing for direct comparison of effect sizes. Shaded area shows standard errors of the coefficients.

Older clades were more likely to have species with vulnerably small range size in the subset of clades with at least 50% of species in the Atlantic Forest, but not in the subsets of all clades or 'Atlantic Forest only' clades. The uncertainty in the statistical effects of age discourages us from attempting further inferences on its effect on the probability of vulnerably small species range size (Table 2).

Discussion

Our results support the hypothesis that the geographic range sizes of the Atlantic Forest's flowering plants are affected by their evolutionary history. Lineages with higher net diversification rate and number of species had, on average, species with smaller range sizes, consistent with our expectation, and distinctly higher probability of vulnerably small range size. Conversely, our evidence does not support a consistent effect of evolutionary age on species range size, although there were indications of a positive effect of clade MRCA age on the probability of vulnerably small range size.

The results are consistent with mechanisms of speciation being a likely underlying cause for present-day range size variation. Clades with a propensity for speciation mechanisms that generate descendants with very small range sizes, e.g. local or budding speciation (Gottlieb 2004, Ferris et al. 2014, Grossenbacher et al. 2014), are likely to show a strong negative correlation between speciation rate and average range size. Such clades may show an even stronger correlation between speciation rate and probability of vulnerably small range size because newly formed species under local speciation (i.e. when small populations diverge at local scale from a wide-range ancestor species) will typically have range sizes smaller than the vulnerability threshold. In contrast, speciation resulting from the split of large portions of the ancestor's range due to vicariance events occurring on large geographical scales may form new species with moderately large ranges from the start, leading to a weaker relationship between speciation rate and probability of vulnerably small range size. In fact, we found that diversification rate relates more strongly to the probability of vulnerably small range than to mean range size, suggesting that speciation mechanisms leading to very small ranges, such as local speciation, may be common in the Atlantic Forest and a driver of species vulnerably small ranges (Levin 1993, Gottlieb 2004, Anacker and Strauss 2014).

Table 2. Coefficients and statistical significance of the effects of clade net diversification rate (under assumptions of negligible extinction $\epsilon = 0$ or high extinction $\epsilon = 0.9$), species richness and MRCA age on the probability that a species has vulnerably small extent of occurrence (n = 1490 species) or area of occupancy (n = 1864 species), according to phylogenetic logistic regression. Data includes all clades with at least 50% of their species occurring in the Atlantic Forest. Vulnerably small extent of occurrence < 20 000 km^2 and area of occupancy < 2000 km^2 were assigned following the threshold of vulnerability adopted in the IUCN Red List of threatened species. Models included life form, habitat and vegetation type as covariates to enable robust evaluation of the independent effects of variables of interest. All quantitative predictors were log transformed, scaled and centred, allowing direct comparison of effect sizes.

		Extent of occurrence			Area of occupancy		
Diversification rate $\epsilon = 0$		0.38***			0.34***		
Diversification rate $\epsilon = 0.9$			0.32***			0.27***	
Species richness				0.41***			0.42***
MRCA age				0.21*			0.12†
Life form							
	Climber	−0.46	−0.22	0.18	−0.10	0.11	0.06
	Herb	0.06	0.35	0.77**	0.39†	0.59**	0.65**
	Shrub	0.43†	0.61*	0.75**	0.58**	0.74***	0.70***
	Tree						
	Unknown/others	0.41	0.70*	1.20***	0.39	0.65*	0.68*
Habitat							
	Aquatic	−0.20	−0.42	−0.54	−0.28	−0.43	−0.35
	Epiphytic	−0.03	−0.19	−0.24	0.24	0.33	0.32
	Hemiepiphyte	−13.57	−13.78	−13.75	−12.27	−12.30	−12.35
	Parasite	1.06	0.82	0.28	0.81	0.66	0.60
	Rupicolous	0.31	0.24	0.13	0.09	0.21	0.08
	Saprophyte	−14.37	−14.51	−14.80	−0.49	−0.53	−0.62
	Terricolous						
	Unknown	−14.24	−14.40	−14.42	0.20	0.07	0.04
Vegetation type							
	Open vegetation	1.26***	1.14***	1.20***	1.50***	1.58***	1.44***
	Forest	1.15***	1.19***	1.17***	1.79***	1.75***	1.83***
	Forest and open veg.						
Intercept		−2.52***	−2.74***	−2.97***	−2.89***	−3.09***	−3.14***

†p < 0.1.
*p < 0.05.
**p < 0.01.
***p < 0.001.

A weak or non-existent effect of age on range size is not unexpected because species 'age' (i.e. divergence time from the most recent common ancestor) does not represent 'time for expansion' of the range. Following a speciation event, the ancestor's range is inherited by the recently diverged species (according to the phylogenetic hypothesis), meaning that the range size of the newly formed species can be already large at age zero and comparable to the ancestor's range. In fact, sister species often show very different range sizes likely due to the asymmetric split of the ancestor's range (Anacker and Strauss 2014). This obscures the link between the time taken to reach the observed range size and the species age (estimated from divergence time), making species age a metric of limited value to test the hypothesis that range size tends to expand with time (Willis 1922). The inconclusive effect of age on range size is also likely affected by the considerable uncertainty in current estimates of MRCA age and species age.

Conservation implications

Species of plant lineages with high diversification rates and large numbers of species are more likely to have vulnerably small range sizes, which is consistent with the hypothesis that those lineages are associated with high species extinction risk (Davies et al. 2011, Tanentzap et al. 2020). This finding illustrates the resilience of plant diversity, because high species' vulnerability to extinction tends to be balanced by high speciation rate and large number of closely related species.

Although many species have small range size due to habitat loss and other human-driven impacts, many others have small range size due to evolutionary reasons. All species with extremely small range size are vulnerable to further habitat loss and associated environmental changes, as well as broader environmental changes (e.g. climate change). However, the reasons for vulnerability are relevant to understanding the contributions of human impacts vs evolutionary dynamics to explain current extinction risk. A species from a plant lineage with extremely high diversification rate and species richness, with many closely related species and naturally occupying a very restricted range, has a vulnerability to extinction largely explained by evolutionary processes. On the other hand, a species that was originally abundant and widespread, and driven to rarity due to overexploitation and habitat loss (e.g. Brazilwood *Paubrasilia echinata*) has a vulnerability largely caused by human impacts. Distinguishing whether each species' vulnerability to extinction is natural or anthropic is likely useful not only to acknowledge distinct conservation stories, but also to inform conservation efforts.

Caveats

Phylogenetic information is improving rapidly and estimates of phylogeny-based variables may change from one study to the other. Age estimates are particularly variable among studies (Stevens 2017), which suggests that both the diversification rate and species age have considerable uncertainty. These uncertainties likely diminish our ability to find age-related patterns, in addition to the problem that species 'age' (i.e. divergence time) is not a good proxy of time for range expansion.

The number of species in a clade can reflect evolutionary diversification rate, but can also reflect bias in species delimitation (Schwartz and Simberloff 2001, Agapow et al. 2004). Criteria for classifying distinct plant individuals or populations – i.e. those with differences in morphology, genetic composition, evolutionary history and/or with isolated reproduction – into multiple or single species may vary across plant families and taxonomists (McDade 1995, Knapp et al. 2005). Some plant lineages present more challenges for definition of clear taxonomic boundaries due to factors such as rates of hybridization, intra-specific variation, convergent evolution and cryptic speciation (McDade 1995, Agapow et al. 2004, Rieseberg and Willis 2007). Whether distinct plant populations are classified into multiple or single species affects species' range size because each of the multiple species usually occupies only a fraction of the range size occupied by a more broadly circumscribed species that encompasses all distinct populations (Brown et al. 1996, Schwartz and Simberloff 2001). Although it is important to be aware of the potential confounding effect of the species delimitation, the most complicated cases where the delimitation is unclear seem to be uncommon (McDade 1995).

Conclusion

By investigating evolutionary patterns in the variation of range size among Atlantic Forest flowering plants, we found that net diversification rate and species richness of lineages markedly affect species' range size. We suggest that the dynamics of speciation and extinction underly variation in range size. Our results illustrate the capacity of plant diversity to persist – because high vulnerability to species extinction tends to be balanced by a large number of closely related species, and help us to recognize the distinct conservation stories of species in clades where vulnerability to extinction is largely explained by evolutionary processes in contrast to species driven to a vulnerable state by overexploitation and habitat loss. Finally, by revealing patterns in the geographic range size of plant species, this study improves our understanding of factors underpinning high levels of plant diversity, rarity and vulnerability to extinction in the Atlantic Forest and likely in other tropical floras.

Data availability statement

Data available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.z34tmpg9v> (Leão et al. 2020).

Acknowledgements – We are particularly grateful to all herbaria that provided plant occurrences through the Species Link network (see list of herbaria in Supplementary material Appendix 1), without which this work would be impossible. We are grateful to members of the Forest Ecology Lab and Jeannine Cavender-Bares Lab, and other colleagues at Univ. of Minnesota for providing insightful comments and criticism, in particular Jose Eduardo Meireles, Jason Reinhardt, George Weiblen, Daniel Stanton and Arindam Banerjee. We are also grateful to colleagues at the Royal Botanic

Gardens – Kew, particularly Barnaby Walker for many constructive discussions, Oscar Alejandro Pérez-Escobar, Rowan Schley and Pablo Hendrigo Alves de Melo.
Funding – Tarciso Leão received Doctoral Fellowship from CAPES no. 1145/12-6, Interdisciplinary Doctoral Fellowship from Univ. of Minnesota and support from the Inst. on the Environment.

References

Agapow, P. M. et al. 2004. The impact of species concept on biodiversity studies. – Quart. Rev. Biol. 79: 161–179.
Anacker, B. L. and Strauss, S. Y. 2014. The geography and ecology of plant speciation: range overlap and niche divergence in sister species. – Proc. R. Soc. B 281: 20132980.
Barber, C. B. et al. 1996. The quickhull algorithm for convex hulls. – ACM Trans. Math. Softw. 22: 469–483.
Bivand, R. et al. 2016. rgdal: bindings for the geospatial data abstraction library. – R package ver. 1.1-10, <https://CRAN.R-project.org/package=rgdal>.
Bivand, R. S. et al. 2013. Applied spatial data analysis with R. – Springer.
Brazil Flora Group 2019. Brazilian Flora 2020 project – Projeto Flora do Brasil 2020, v393.179. – Inst. de Pesquisas Jardim Botanico do Rio de Janeiro.
Brown, J. H. et al. 1996. The geographic range: size, shape, boundaries, and internal structure. – Annu. Rev. Ecol. Syst. 27: 597–623.
Brummitt, N. et al. 2015. The sampled red list index for plants, phase II: ground-truthing specimen-based conservation assessments. – Phil. Trans. R. Soc. B 370: 20140015.
Carvalho, G. 2016. flora: tools for interacting with the Brazilian flora 2020. – R package ver. 0.2.7, <https://CRAN.R-project.org/package=flora>.
Castiglione, S. et al. 2017. Diversification rates and the evolution of species range size frequency distribution. – Front. Ecol. Evol. 5: 1–10.
Davies, T. J. et al. 2011. Extinction risk and diversification are linked in a plant biodiversity hotspot. – PLoS Biol. 9: e1000620.
Ferris, K. G. et al. 2014. Speciation on a local geographic scale: the evolution of a rare rock outcrop specialist in *Mimulus*. – Phil. Trans. R. Soc. B 369: 20140001.
Fox, J. and Weisberg, S. 2011. An R companion to applied regression. – SAGE Publications.
Gallagher, R. V. 2016. Correlates of range size variation in the Australian seed-plant flora. – J. Biogeogr. 43: 1287–1298.
Gaston, K. J. 2003. The structure and dynamics of geographic ranges. – Oxford Univ. Press.
Gaston, K. J. and Fuller, R. A. 2009. The sizes of species' geographic ranges. – J. Appl. Ecol. 46: 1–9.
Gottlieb, L. D. 2004. Rethinking classic examples of recent speciation in plants. – New Phytol. 161: 71–82.
Grossenbacher, D. L. et al. 2014. Niche and range size patterns suggest that speciation begins in small, ecologically diverged populations in North American monkeyflowers (*Mimulus* spp.). – Evolution 68: 1270–1280.
Habel, K. et al. 2015. geometry: mesh generation and surface tesselation. – R package ver. 0.3-6, <https://CRAN.R-project.org/package=geometry>.
Harmon, L. J. 2019. Phylogenetic comparative methods. – Independent Publisher.
Hijmans, R. J. 2016. raster: geographic data analysis and modeling. – R package ver. 2.5-8, <https://CRAN.R-project.org/package=raster>.
IUCN 2012. IUCN red list categories and criteria: version 3.1, 2nd ed. – IUCN.
Ives, A. R. 2018. R2s for correlated data: phylogenetic models, LMMs, and GLMMs. – Syst. Biol. 68: 234–251.
Ives, A. R. and Garland, T. 2010. Phylogenetic logistic regression for binary dependent variables. – Syst. Biol. 59: 9–26.
Knapp, S. et al. 2005. Taxonomic inflation, species concepts and global species lists. – Trends Ecol. Evol. 20: 7–8; author reply 8–9.
Leão, T. C. C. et al. 2014. Predicting extinction risk of Brazilian Atlantic forest angiosperms. – Conserv. Biol. 28: 1349–1359.
Leão, T. C. C. et al. 2020. Data from: Evolutionary patterns in the geographic range size of Atlantic Forest plants. – Dryad Digital Repository. doi:10.5061/dryad.z34tmpg9v.
Levin, D. A. 1993. Local speciation in plants: the rule not the exception. – Syst. Bot. 18: 197–208.
Lozano, F. D. and Schwartz, M. W. 2005. Patterns of rarity and taxonomic group size in plants. – Biol. Conserv. 126: 146–154.
Magallon, S. and Sanderson, M. J. 2001. Absolute diversification rates in angiosperm clades. – Evolution 55: 1762–1780.
Magallón, S. et al. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. – New Phytol. 207: 437–453.
Martinelli, G. and Moraes, M. A. 2013. Livro vermelho da flora do Brasil. – Inst. de Pesquisas Jardim Botânico do Rio de Janeiro.
Martins, E. et al. 2018. Brazilian efforts towards achieving a comprehensive extinction risk assessment for its known flora. – Rodriguésia 69: 1529–1537.
McDade, L. A. 1995. Species concepts and problems in practice: insight from botanical monographs. – Syst. Bot. 20: 606–622.
Morueta-Holme, N. et al. 2013. Habitat area and climate stability determine geographical variation in plant species range sizes. – Ecol. Lett. 16: 1446–1454.
Murray, B. R. et al. 2002. How plant life-history and ecological traits relate to species rarity and commonness at varying spatial scales. – Austral Ecol. 27: 291–310.
Nic Lughadha, E. et al. 2019. The use and misuse of herbarium specimens in evaluating plant extinction risks. – Phil. Trans. R. Soc. B 374: 20170402.
Pagel, M. 1999. Inferring the historical patterns of biological evolution. – Nature 401: 877–884.
Paradis, E. et al. 2004. APE: analyses of phylogenetics and evolution in R language. – Bioinformatics 20: 289–290.
Paul, J. R. et al. 2009. Evolutionary time for dispersal limits the extent but not the occupancy of species' potential ranges in the tropical plant genus *Psychotria* (Rubiaceae). – Am. Nat. 173: 188–199.
Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). – Methods Ecol. Evol. 3: 217–223.
Ribeiro, M. C. et al. 2009. The Brazilian Atlantic Forest: how much is left, and how is the remaining forest distributed? Implications for conservation. – Biol. Conserv. 142: 1141–1153.
Rieseberg, L. H. and Willis, J. H. 2007. Plant speciation. – Science 317: 910–914.
Rosenzweig, M. L. 2001. Loss of speciation rate will impoverish future diversity. – Proc. Natl Acad. Sci. USA 98: 5404–5410.
Scarano, F. R. and Ceotto, P. 2015. Brazilian Atlantic forest: impact, vulnerability, and adaptation to climate change. – Biodivers. Conserv. 24: 2319–2331.

Schwartz, M. W. and Simberloff 2001. Taxon size predicts rates of rarity in vascular plants. – Ecol. Lett. 4: 464–469.
Smith, S. A. and Brown, J. W. 2018. Constructing a broadly inclusive seed plant phylogeny. – Am. J. Bot. 105: 302–314.
Stevens, P. F. 2017. Angiosperm phylogeny website, version 13. – <www.mobot.org/MOBOT/research/APweb>.
Tanentzap, A. J. et al. 2020. Does evolutionary history correlate with contemporary extinction risk by influencing range size dynamics? – Am. Nat. 195: 569–576.
The Brazil Flora Group et al. 2018. Brazilian flora 2020: innovation and collaboration to meet target 1 of the Global Strategy for Plant Conservation (GSPC). – Rodriguésia 69: 1513–1527.

Supplementary material (available online as Appendix ECOG-05160 at <www.ecography.org/appendix/ecog-05160>). Appendix 1.

Title, P. O. and Rabosky, D. L. 2019. Tip rates, phylogenies, and diversification: what are we estimating, and how good are the estimates? – Methods Ecol. Evol. 10: 821–834.
Tung Ho, L. S. and Ané, C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. – Syst. Biol. 63: 397–408.
Willis, F. et al. 2003. Defining a role for herbarium data in Red List assessments: a case study of *Plectranthus* from eastern and southern tropical Africa. – Biodivers. Conserv. 12: 1537–1552.
Willis, J. C. 1922. Age and area: a study in geographical distribution and origin of species. – Cambridge Univ. Press.
Zappi, D. C. et al. 2015. Growing knowledge: an overview of seed plant diversity in Brazil. – Rodriguésia 66: 1085–1113.